

CLAIMS

We claim:

1. A method comprising:
 - accepting a string of characters representing a possibly incorrectly entered URL;
 - parsing the string into a set of URL parts, a part formed from characters having values in a first space of characters, each part having a corresponding distance measure of closeness for measuring distances between URL parts;
 - forming a signature of each URL part, forming said signature including transforming the characters of the URL part whose values are in the first space into characters in a second space such that the distance measure of closeness is transformed to a distance measure of closeness that is not necessarily integer valued;
 - for each URL part, searching for at least one cluster of a set of pre-formed clusters, the set of pre-formed clusters being clusters of valid URL parts that are close according to the distance measure of closeness that is not necessarily integer valued, each cluster in the set of pre-formed clusters having a representative URL part and signature thereof, the searching using the signature of the URL part;
 - further searching for a valid URL part within each cluster found in the searching step.
2. A method for resolving a possibly incorrectly entered URL comprising:
 - accepting the entered URL;
 - parsing the accepted URL into URL parts;
 - carrying out a conventional URL lookup; and
 - for any part of the accepted URL that is not valid:
 - determining a signature for the accepted URL part; and
 - conducting a fuzzy search for at least one valid URL part that is close to the invalid URL part according to a distance measure that

combines at least one local measure, each measure suited for a particular type of URL part.

3. A method as recited in claim 2, wherein the not-valid accepted URL part includes characters in a first space wherein a distance measure of closeness is integer-valued, and wherein the determining of the signature of the accepted URL part includes converting the first space into a second space such that the signature of the URL part is a sequence of values in the second space, the second space being a space wherein the distance measure for comparing signatures of URL parts is non-integer or a general distance function in a metric space such that cluster analysis can be performed on signatures of valid URLs or URL parts.
4. A method as recited in claim 2, further comprising forming at least one valid URLs from the URL parts found in the fuzzy search.
5. A method as recited in claim 2, wherein each URL part includes characters from a corresponding alphabet, the alphabet being of letters in the case of text, of phonemes in the case of sounds, of glyphs in the case of glyphs/images, and of numbers in the case of a numerical URL parts.
6. A method as recited in claim 2, wherein conducting the fuzzy search occurs at different parts of a computer network according to the type of part that was carrying out the conventional URL lookup determines is not valid.
7. A method as recited in claim 2, wherein conducting the fuzzy search includes:
 - determining at least one cluster of a set of pre-formed clusters wherein the accepted URL part is likely to be, each cluster comprising a set of valid URL parts that are close according to a distance measure and having a representative URL part having a known signature, the determining including finding the at least one signatures of representative URLs close to the signature of the accepted URL part; and
 - further searching for a valid URL part within the at least one determined clusters.

8. A method as recited in claim 2, wherein the not-valid accepted URL part includes characters in a first space wherein a distance measure of closeness is integer-valued, and wherein the determining of the signature of the accepted URL part includes converting the first space into a second space such that the signature of the URL part is a sequence of values in the second space, the second space being a space wherein the distance measure for comparing signatures of URL parts is non-integer such that cluster analysis can be performed on signatures of valid URLs or URL parts.
9. A method as recited in claim 7, wherein the not-valid accepted URL part includes characters in a first space wherein a distance measure of closeness is integer-valued, and wherein the determining of the signature of the accepted URL part includes converting the first space into a second space such that the signature of the parts URL is a sequence of values in the second space, the second space being a space wherein the distance measure for comparing signatures of URL parts is non-integer such that cluster analysis can be performed on signatures of valid URLs or URL parts.
10. A method as recited in claim 8, wherein the second space is n-dimensional Euclidean space such that the signature of the URL part is a sequence of n-dimensional vectors.
11. A method as recited in claim 10, wherein the values of the converted characters are on the unit sphere such that the second space is transformed to the n-dimensional unit sphere, wherein the signature of the URL part is a sequence of n-dimensional vectors on the unit sphere, and wherein calculating the distance between two URL parts can be carried out by a convolution-like operation on the signatures of the two URL parts.
12. A method as recited in claim 2, wherein the local measures include at least one distance measures from the set consisting of a string comparison measure for URL parts for which string comparison is appropriate, a phonetic difference for URL parts for which sound comparison is appropriate, image comparison for URL parts that include glyphs/images for which image comparison is appropriate, and a numerical difference for URL parts that are numerical and for which numerical comparison is appropriate.
13. A method as recited in claim 6, wherein the computer network is the Internet.
14. A method as recited in claim 6, wherein the computer network is a private network.

15. A method as recited in claim 4 further including the steps of:
 - displaying a list of formed URLs to a user of a IP client; and
 - prompting the user of the IP client to select a one of formed URLs in the list.
16. A method as recited in claim 2 wherein the fuzzy search is performed at a client of a computer network.
17. A method as recited in claim 2 wherein the fuzzy search is performed at a server of a computer network.
18. A carrier medium carrying at least one computer readable code segment for instructing a processor of a processing system to implement a method for resolving a possibly incorrectly entered URL, the method comprising:
 - accepting the entered URL;
 - parsing the accepted URL into URL parts;
 - carrying out a conventional URL lookup; and
 - for any part of the accepted URL that is not valid:
 - determining a signature for the accepted URL part; and
 - conducting a fuzzy search at least one valid URL part that is close to the invalid URL part according to a distance measure that combines at least one local measure, each measure suited for a particular type of URL part.
19. A carrier medium as recited in claim 18, wherein the method further comprises forming at least one valid URL from the URL parts found in the fuzzy search.
20. A carrier medium as recited in claim 18, wherein conducting the fuzzy search includes:
 - determining at least one cluster of a set of pre-formed clusters wherein the accepted URL part is likely to be, each cluster comprising a set of valid URL parts that are close according to a distance measure and having a representative URL

part having a known signature, the determining including finding the at least one signature of at least one representative URL close to the signature of the accepted URL part; and

further searching for a valid URL part within the at least one determined cluster.

21. A carrier medium as recited in claim 18, wherein the not-valid accepted URL part includes characters in a first space wherein a distance measure of closeness is integer-valued, and wherein the determining of the signature of the accepted URL part includes converting the first space into a second space such that the signature of the URL part is a sequence of values in the second space, the second space being a space wherein the distance measure for comparing signatures of URL parts is non-integer such that cluster analysis can be performed on signatures of valid URLs or URL parts.
22. A carrier medium as recited in claim 20, wherein the not-valid accepted URL part includes characters in a first space wherein a distance measure of closeness is integer-valued, and wherein the determining of the signature of the accepted URL part includes converting the first space into a second space such that the signature of the parts URL is a sequence of values in the second space, the second space being a space wherein the distance measure for comparing signatures of URL parts is non-integer such that cluster analysis can be performed on signatures of valid URLs or URL parts.
23. A carrier medium as recited in claim 21, wherein the second space is n-dimensional Euclidean space such that the signature of the URL part is a sequence of n-dimensional vectors.
24. A carrier medium as recited in claim 23, wherein the values of the converted characters are on the unit sphere such that the second space is transformed to the n-dimensional unit sphere, wherein the signature of the URL part is a sequence of n-dimensional vectors on the unit sphere, and wherein calculating the distance between two URL parts can be carried out by a convolution-like operation on the signatures of the two URL parts.

25. A carrier medium as recited in claim 18, wherein the local measures include at least one distance measure from the set consisting of a string comparison measure for URL parts for which string comparison is appropriate, a phonetic difference for URL parts for which sound comparison is appropriate, and a numerical difference for URL parts that are numerical and for which numerical comparison is appropriate.
26. A carrier medium as recited in claim 18, wherein the local measures include one or more distance measures from the set consisting of a string comparison measure for URL parts for which string comparison is appropriate, a phonetic distance for URL parts for which sound comparison is appropriate, a glyph distance between images/icons for parts where it is appropriate, and a numerical distance for URL parts that are numerical and for which numerical comparison is appropriate.
27. A method of conducting the fuzzy search for a source URL part that closely matches a valid URL part, comprising:
 - determining a signature for the source URL part;
 - determining at least one cluster of a set of pre-formed clusters wherein the source URL part is likely to be, each cluster comprising a set of valid URL parts that are close according to a distance measure and having a representative URL part having a known signature, the determining of the likely clusters including finding the at least one signature of representative URLs close to the signature of the accepted URL part; and
 - further searching for a valid URL part within the at least one determined cluster.
28. A method as recited in claim 27, wherein the source URL part includes characters in a first space wherein a distance measure of closeness is integer-valued, and wherein the determining of the signature of the source URL part includes converting the first space into a second space such that the signature of the URL part is a sequence of values in the second space, the second space being a space wherein the distance measure for comparing signatures of URL parts is non-integer such that cluster analysis can be performed on signatures of valid URLs or URL parts.

29. A method as recited in claim 27, wherein the source URL part includes characters in a first space wherein a distance measure of closeness is integer-valued, and wherein the determining of the signature of the source URL part includes converting the first space into a second space such that the signature of the URL part is a sequence of values in the second space, the second space being a space wherein the distance measure for comparing signatures of URL parts is non-integer but a general distance function in a metric space, such that cluster analysis can be performed on signatures of valid URLs or URL parts.

30. A method as recited in claim 28, wherein the second space is n-dimensional Euclidean space such that the signature of the URL part is a sequence of n-dimensional vectors.

31. A method as recited in claim 30, wherein the values of the converted characters are on the unit sphere such that the second space is transformed to the n-dimensional unit sphere, wherein the signature of the URL part is a sequence of n-dimensional vectors on the unit sphere, and wherein calculating the distance between two URL parts can be carried out by a convolution-like operation on the signatures of the two URL parts.

32. A method comprising:

- accepting a string of symbols representing a possibly incorrectly entered URL, each symbol being a text character, phoneme or glyph;
- parsing the string into a set of URL parts, a part formed from characters having values in a first space of characters, each part having a corresponding distance measure of closeness for measuring distances between URL parts;
- forming a signature of each URL part, forming said signature including transforming the characters of the URL part whose values are in the first space into characters in a second space such that the distance measure of closeness is transformed to a distance measure of closeness that is not necessarily integer valued;
- for each URL part, searching for one or more clusters of a set of pre-formed clusters, the set of pre-formed clusters being clusters of valid URL parts that are close according to the distance measure of closeness that is not necessarily integer valued, but a general distance function in a metric space, each cluster in the set of pre-formed clusters having a

representative URL part and signature thereof, the searching using the signature of the URL part;

further searching for a valid URL part within each cluster found in the searching step.